Silicon Meets Compound Semiconductors: Pioneering Wireless Communications

N. Collaert[1], R. Alcotte, A. Alian, M. Asad, I. Bagal[2], S. Banerjee, G. Boccardi, P. Cardinael[3], I. Comart, C. Desset, R. ElKashlan, F. Filice, G. Gramegna, H. Jafarpoorchekab, A. Khaled, A. Kumar, B. Kunert, Y. Mols, B. O'Sullivan, S. Park, U. Peralagu, N. Pinho, A. Rathi, A. Sibaja-Hernandez, S. Sinha, D. Smellie[2], X. Sun, A. Vais, B. Vanhouche[1], B. Vermeersch, D. Xiao, S. Yadav, D. Yan[1], H. Yu, Y. Zhang, J. Van Driessche, P. Wambacq[1], M. Peeters, B. Parvais[1]

Imec, Heverlee, Belgium; [1]Also with VUB, Belgium; [2]Also with KU Leuven, Belgium; [3]Also with UC Louvain, Belgium.

Keywords: GaN, InP, RF interposer, heterogeneous integration

Abstract

The rapid evolution of wireless communication requires advanced semiconductors for 5G(+) and upcoming 6G networks. GaN and InP stand out, enhancing RF Front-End performance. This paper highlights their benefits in high-frequency applications for 5G(+) and 6G, exploring cost-efficient implementations on Si. It also discusses heterogeneous integration with CMOS, paving the way for the next-gen wireless technologies.

INTRODUCTION

To meet the growing demand for faster data rates, lower energy consumption, and reduced latency, the wireless communication industry is increasing its operating frequencies. The focus is currently on the FR3 spectrum (6-24GHz), where GaN technology is expected to play a crucial role. Despite delays in millimeter-wave deployment, it is anticipated to debut around 2026. Looking ahead, 6G technology is slated for introduction post-2030 to address the expanding bandwidth demands, with discussions even exploring frequencies beyond 100GHz.

Secure long-distance links at these frequencies require beamforming, facilitated by reduced wavelength and the ability to pack more antennas into a confined space. This, in turn, alleviates performance requirements for power amplifiers (PAs). However, challenges include decreased output power and lower efficiency, especially in mm-wave and sub-THz frequencies.

While CMOS is favored for its cost and integration advantages, compound semiconductors like GaN and InP are superior due to their higher mobility, charge density, and breakdown voltage. These devices are typically produced on smaller, costlier substrates, using processes like e-beam and Au metallization, less suitable for high-volume Si-centric applications.

This work will delve into the challenges and benefits associated with scaling up these compound semiconductor technologies. It will also address the necessity of heterogeneous integration, particularly in combining GaN or InP with CMOS, and explore various options for achieving this integration.

POWER AND EFFICIENCY

We evaluated the energy efficiency impact of technology choices on the transmitter (TX) and receiver (RX) using a power model [1, 2]. Fig. 1 indicates that the TX side, particularly the power amplifiers (PAs), dominates the power consumption, especially at longer distances and in footprint-restricted scenarios. This particular study was done for 140GHz, but the same holds true for lower operating frequencies.

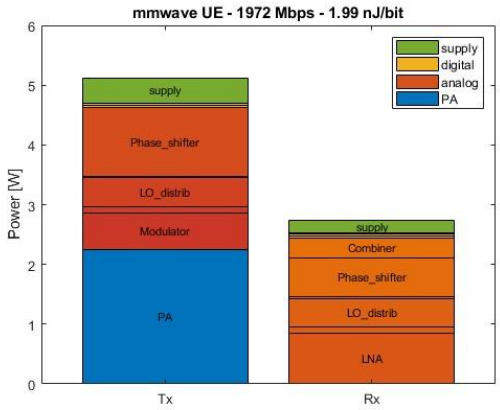

Fig. 1. Breakdown of power consumption of a CMOS TRX with 48 antennas, 4.5dBm per PA, operating at 140GHz and delivering 1.9Gbps at 1.99nJ/bit power efficiency [1].

GaN and InP have emerged as compelling choices for RF front-end modules because of their unique material properties and performance advantages, addressing mm-wave and sub-THz challenges. However, transitioning from CMOS to these technologies requires innovative approaches, materials, and manufacturing techniques to compete in RF applications. The integration challenge due to lattice and thermal expansion mismatch necessitates ongoing research into packaging-based solutions, as described in the following sections.

GAN-SI DEVICES

Riding the wave of GaN power and µLED popularity, RF GaN has emerged as a compelling choice for the PAs in RF Front-End Modules (RF-FEM) due to its exceptional performance. GaN's high electron mobility, charge density and breakdown voltage make it a well-suited candidate for efficient power amplification. Of particular note is the more cost-effective GaN-Si technology, which can be scaled up to 8-inch and 12-inch sizes, generating significant interest today. Within the context of FR3, RF GaN-Si is being considered for two use cases: infrastructure and handset applications, next to applications in SATCOM and aerospace. Fig. 2 shows a TEM picture of a multi-finger RF device with a 3-layer Cu BEOL.

GaN-Si technology, ideal for infrastructure applications with voltages up to 28V and 48V, emerges as a cost-effective and sustainable alternative to GaN-SiC, particularly in base stations for wireless communication networks. However, efficient thermal management is imperative due to the heat generated by GaN-based power amplifiers, crucial for maintaining optimal performance and reliability. In addition, the evaluation of GaN-Si for handset applications presents challenges in achieving a good E-mode device at V_{DD} < 6V, necessitating the introduction of new modules and materials.

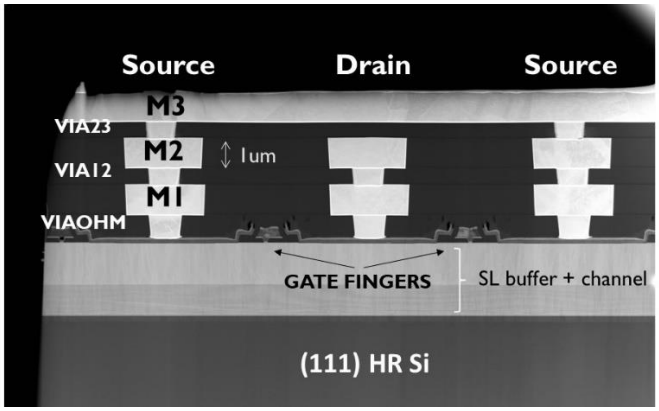

Fig. 2. TEM image of a multi-finger RF device with 3 levels of BEOL metal. A superlattice (SL) buffer is grown on top of (111) high resistivity (HR) silicon.

The introduction of S/D regrowth to reduce the access resistance and gate dielectrics to reduce the gate leakage are two topics being studied in that context. Benchmarking against GaN-SiC at 6-18GHz reveals promising results, but ongoing efforts focus on enhancing V_{DD} and output power without sacrificing efficiency for infrastructure applications and addressing the E-mode challenges in handsets (Fig. 3).

To comprehensively grasp the trade-offs inherent in device design and discern their impact at the circuit level, the incorporation of physics-based models becomes imperative. These models must effectively account for non-idealities arising from various sources, including the increased defect density [13], diminished thermal conductivity in comparison to GaN-SiC [14], and the substrate losses [15].

In a comprehensive study detailed in [16], the prerequisites for these models are thoroughly examined, forming the basis for establishing a Design-Technology Co-Optimization (DTCO) loop. This iterative loop is designed to provide crucial insights into whether non-idealities can be effectively managed at the circuit and system level. Techniques such as digital pre-distortion (DPD) come into play, as they are evaluated for their potential to mitigate the impact of non-idealities. Simultaneously, the analysis aims to ascertain whether breakthroughs at the material and device levels are also essential for reaching the best performance.

Recent studies also look into expanding the capabilities of GaN pushing its performance boundaries to accommodate operating frequencies as high as 94GHz and beyond [17]. In [18], devices with f_{max} as high as 700GHz have been presented. These endeavors include innovative approaches at the device design stage, exemplified by the use of N-polar GaN [19]. Such advancements showcase a holistic perspective, considering not only the mitigation of non-idealities but also the enhancement of intrinsic device characteristics to cater to the demands of emerging high-frequency applications. As the industry continues to explore these frontiers, the intersection of material innovations, device design strategies, and circuit-level optimizations becomes increasingly crucial in shaping the trajectory of these semiconductor technologies.

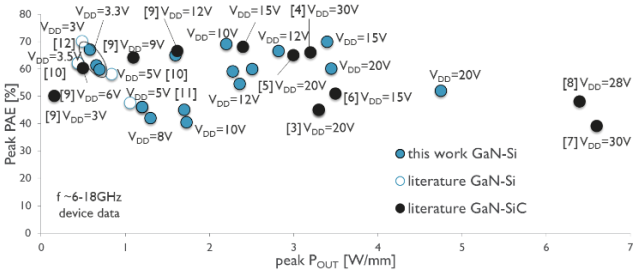

Fig. 3. Power Added Efficiency (PAE) plotted against Output Power (Pout) for GaN-Si and GaN-SiC devices, obtained from load-pull measurements within the 6-18GHz frequency range at varying V_{DD} levels.

INP-SI DEVICES

In Fig. 4, InP HBT technology is shown to excel at 140GHz, offering the best efficiency-power trade-off among Si(Ge) (Bi)CMOS, Si FinFET and RFSOI technologies. However, InP devices, whether HBT or HEMT, lag behind GaN, GaAs, and CMOS due to limited commercialization, primarily stemming from high production costs and the use of these technologies in niche markets.

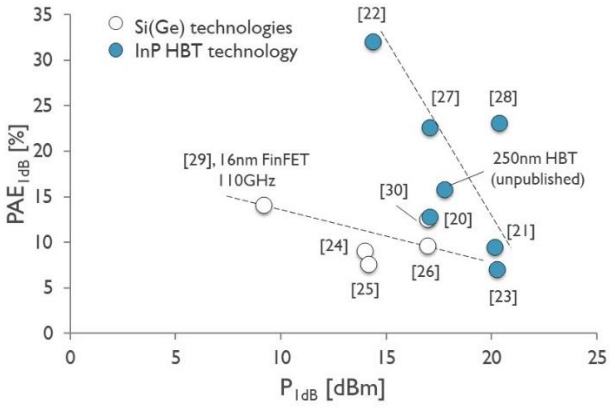

Fig. 4. Power Added Efficiency at 1 dB compression (PAE1db) compared to 1 dB Compression Output Power (P1db) for power amplifiers designed in SiGe BiCMOS, Si FinFET, RFSOI, and InP HBT technologies at 140GHz; the trendlines shown are a guidance to the eye.

Achieving the essential upscaling, maturation, and enhanced compatibility of InP technologies with CMOS and heterogeneous integration is imperative. To address these challenges, various strategies have been proposed. These encompass techniques such as direct growth of III-V materials on silicon, including nano-ridge engineering (NRE) for defect reduction [31-33], wafer reconstruction [34, 35], and micro-transfer printing [36]. Wafer reconstruction involves the transfer of InP tiles, with or without active layers, onto larger silicon wafers through die-to-wafer bonding or layer transfer. Subsequently, this reconstructed wafer can undergo further processing for device fabrication. The main advantage is that one can start with high crystal quality. Micro-transfer printing is a collective die-to-wafer approach wherein (partially) processed dies are transferred onto the target wafer using a stamp. The major difference is that in this case the dies can be (pre-)processed including some layers of metallization, and the method of layer transfer is different. Typically post-processing of this wafer entails metallization, connecting the different dissimilar devices on the wafer. In general, we can say that there are different ways of fabricating these reconstructed wafers, and this might be very tailored to the specific application. Today's challenges lie in ensuring the effective transfer of materials and the removal of the InP substrate. Numerous techniques are currently under consideration including grinding, laser lift-off, and SmartCut™ [35]. Next to that, managing warpage induced by the coefficient of thermal expansion (CTE) mismatch among the different materials in the reconstructed assembly, addressing issues like delamination and ensuring proper planarization, remains also an important topic of research and development.

Ultimately, the success of InP-on-Si technology hinges on achieving excellent device performance and cost-effectiveness. Both the direct growth and wafer reconstruction methods offer advantages over small InP substrates, but further cost reductions in III-V growth and InP substrates are essential for these approaches. Next to that, sustainability and environmental considerations play a crucial role in down-selecting the different InP-Si approaches. As the demand for high-performance electronic and optoelectronic devices continues to rise, it becomes imperative to balance technological advancements with sustainability metrics like energy efficiency, resource conservation, and reduced environmental impact.

HETEROGENEOUS INTEGRATION

In the quest for optimal power and efficiency, we will need to combine CMOS ICs (for calibration and control) with compound semiconductor technologies such as GaN and InP. The size of the antenna array is dependent on the square of the wavelength (λ^2) [37] where frequencies exceeding 70GHz result in the antenna module defining the transceiver's available area. To address this challenge, innovative solutions like 3D integration and interposer technologies (chiplets) are being explored, facilitating compact and high-performance designs with efficient connections, especially between the antenna and RF-FEM.

However, thermal management remains a significant concern in 3D integration, necessitating effective heatsink solutions through thermal vias or direct die-to-heatsink contact. In mobile handsets, where antenna numbers can be reduced, 2.5D and wafer-level interposer technologies offer new possibilities. In [38], an RF-tailored Si interposer is introduced. This work presents a highly scaled RF interposer platform (300mm) on low-resistivity Si (15-25 Ωcm) for heterogenous integration in high-frequency applications. A schematic presentation of the RF interposer is shown in Fig. 5 and compared to a typical high-speed digital interposer. The platform exhibits interconnect insertion loss below 0.3 dB/mm at 100 GHz (as measured), with integrated passives like inductors. Various RF interposer inductors have been characterized, demonstrating improved Q-factors and resonance frequencies as compared to on-chip solutions. Inductors with Q_{max} above 40 have been demonstrated.

One of the key features of this platform is the use of standard low-resistivity Si substrates. This eliminates the need for the use of more expensive high resistivity substrates where one of the challenges is to enable proper surface passivation. Specific field plates have been included in the bottom digital-like metal lines in order to reduce the substrate losses. Next to that, the platform also enables flip-chip performance up to 500 GHz (as confirmed by simulations), facilitating heterogenous integration of mm-wave ICs in different technologies.

To achieve this performance, the RF Si interposer includes an 18 µm thick benzocyclobutene (BCB) layer and 5 µm Cu RDL to minimize parasitic capacitance and transmission line loss. The narrow pitch of µbumps allows for efficient flip-chip performance, surpassing traditional PCB technologies. Fig. 6 shows a Si die µ-bumped to the RF interposer substrate.

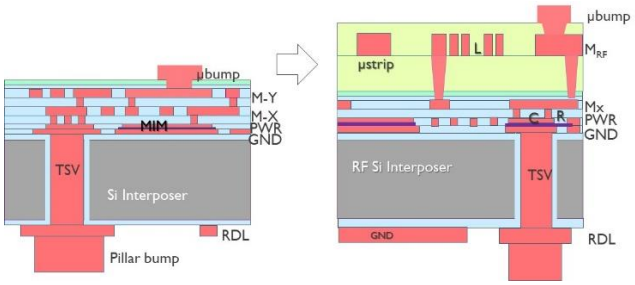

Fig. 5. Schematic presentation of the RF interposer (right) and comparison to the typical digital interposer (left); Mx=fine-grained metal layers, MRF=metal layers for RF signal routing, RDL=redistribution layers, TSV=through silicon vias, MIM=metal-insulator-metal capacitor, PWR=power line and GND=ground.

This cost-effective RF Si interposer on low-resistivity Si substrates offers high-Q passives, low-loss interconnects, and flip-chip capabilities up to 500 GHz. The platform's superior RF performance and thermal management make it suitable for heterogenous integration of mm-wave ICs in various technologies, enabling co-design of package, RFIC, and antenna for beyond-5G applications.

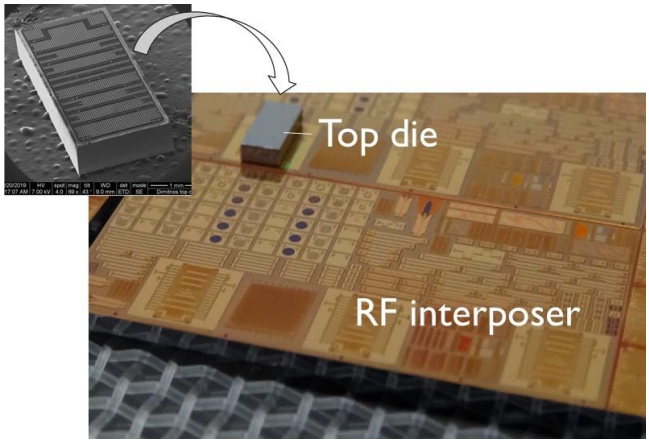

Fig. 6. Top view of a stacked (passive) Si die µ-bumped on the RF interposer [38].

SUMMARY

In conclusion, the integration of GaN and InP technologies with silicon holds immense promise for advancing the capabilities of RF-FEM in wireless communication systems. As the demand for higher frequencies and power efficiency continues to grow with the evolution from 5G(+) to 6G networks, overcoming the challenges of scaling and cost-effectiveness becomes paramount. Heterogeneous integration, combining the strengths of compound semiconductors and CMOS, emerges as a key strategy to unlock the full potential of these technologies, paving the way for the next era in wireless communication. In this work, we have briefly focused on the enablement of an RF-tailored interposer using standard Si substrate and the use of BCB dielectrics combined with copper RDL to address challenges in RF performance. It is one of the key building blocks to enable advanced heterogeneous RF systems addressing a variety of different future applications.

ACKNOWLEDGEMENTS

The contributions of the Advanced RF and 3D system integration research teams at imec are acknowledged. This work is supported by imec's IIAP program members and partially supported by the Horizon Europe project ICOS (Grant Agreement no. 101092562).

REFERENCES

[1] C. Desset et al., PIMRC Workshop, UK, Aug. 2020
[2] C. Desset et al., GlobeCom Workshop, 2021
[3] https://gems.ece.gatech.edu/PA_survey.html,
[4] S. Piotrowicz et al., IMS 2014
[5] M. Marchetti et al., COMCAS 2015
[6] A. Arias et al., CSICS 2017
[7] K. Kiziloglu et al., IEEE MTT-S International Microwave Symposium Digest 1999
[8] O. Jardel et al., EuMA 2010
[9] M.-H. Weng et al., 2013
[10] Y. Zhou et al., J-EDS 2021
[11] H. Xie et al., Applied Physics Express 2021
[12] H. Xie et al., IEEE Microwave and Wireless Components Letters 2021
[13] V. Putcha et al., IRPS 2021
[14] B. Vermeersch et al., IEDM 2022 (invited)
[15] P. Cardinael et al., IEEE MTT-S, 2022
[16] B. Parvais et al., SISPAD 2021
[17] H.-W. Then et al., IEDM 2021
[18] H. W. Then et al., IEEE Microwave and Wireless Technology Letters, vol. 33, no. 6, June 2023
[19] W. Liu et al., IEEE Microwave and Wireless Components Letters 2021
[20] A. S. H. Ahmed et al., MTT-S 2020
[21] A. S. H. Ahmed et al., EUMIC 2020
[22] K. Ning et al., RFIC 2020
[23] Z. Griffith et al., MTT-S, 2019
[24] A. Visweswaran et al., ESSCIRC 2019
[25] S. Li et al., RFIC 2021
[26] X. Li et al., 19.1, ISSCC 2022
[27] A. Banerjee et al., IRMMW-TH 2022
[28] S. Gielen et al., RFIC 2023
[29] Q. Yu et al., RFIC 2022
[30] X. Li et al., JSSC 2022
[31] B. Kunert et al., Compound Semiconductor 2018
[32] A. Vais et al., IEDM 2019
[33] S. Yadav et al., EUMIC 2020
[34] S. Manda et al, IEDM 2019
[35] B. Ghyselen et al., Physica Status Solidi (A) Applications and Materials, 2021
[36] E. Soltanian et al., IEEE Silicon Photonics Conference (SiPhotonics), 2023
[37] B. Sadhu et al., IEEE Microwave magazine, 2019
[38] X. Sun et al., ECTC, 2022
[39] S. Sinha et al., IEEE Microwave and Wireless Technology Letters, 2023